OXFORD

Quantile regression estimands and models: revisiting the motherhood wage penalty debate

Nicolai T. Borgen,[1,2,*] **ⓘD Andreas Haupt**[3,] **ⓘD and Øyvind Nicolay Wiborg**[2,] **ⓘD**

[1]Department of Special Needs Education, University of Oslo, Norway
[2]Department of Sociology and Human Geography, University of Oslo, Norway
[3]Institute of Sociology, Media and Cultural Studies, Karlsruhe Institute of Technology, Germany

*Corresponding author. Email: n.t.borgen@isp.uio.no

Abstract

This paper discusses the crucial but sometimes neglected differences between unconditional quantile regression (UQR) models and quantile treatment effects (QTE) models. We argue that there is a frequent mismatch between the aim of the quantile regression analysis and the quantitative toolkit used in much of the applied literature, including the motherhood wage penalty literature. This mismatch may result in wrong conclusions being drawn from the data, and in the end, misguided theories. In this paper, we clarify the crucial conceptual distinction between influences on quantiles of the overall distributions, which we term population-level influences, and individual-level QTEs. Further, we use data simulations to illustrate that various classes of quantile regression models may, in some instances, give entirely different conclusions (to different questions). Finally, we compare quantile regression estimates using real data examples, showing that UQR and QTE models differ sometimes but not always. Still, the conceptual and empirical distinctions between quantile regression models underline the need to match the correct model to the specific research questions. We conclude the paper with a few practical guidelines for researchers.

Introduction

The development of the counterfactual model of causality alongside new statistical methods, enhanced computational power, continuous software development, and increasing data availability has allowed for exploring new research questions and getting better answers to old ones. At the same time, the complexity of some of these models makes them vulnerable to misinterpretation and misuse. One prime example is the exciting innovations of the quantile regression framework that has taken place over the last decade. The development of the unconditional quantile regression (UQR) model (Firpo, Fortin and Lemieux, 2009) and the somewhat less-known quantile treatment effects (QTE) models (Firpo, 2007; Powell, 2020; Borgen, Haupt and Wiborg, 2021b) has the potential for novel findings. However, the complexity of quantile regression models has made them poorly understood in much of the applied social science literature (Wenz, 2019; Rios-Avila and Maroto, 2020).

This paper focuses on the crucial differences between UQR and QTE models. We believe that there is a frequent mismatch between the aim of the quantile regression analysis and the quantitative toolkit used in much of the applied literature; quantile treatment effects have gained popularity as estimand on a theoretical level since the 2000s, whereas unconditional quantile regression has been on the rise on a methodological level. As an illustration, consider the recent debate on the motherhood wage penalty, published in top-rated journals such as American Sociological Review (Killewald and Bearak, 2014; England et al., 2016), European Sociological Review (Cooke, 2014), and Demography (Glauber, 2018). Whether the motherhood disadvantage increases or decreases across the wage distribution is a question of quantile treatment effects. Yet, studies have used UQR to examine these wage gaps rather than QTE models.

Regarding the motherhood wage penalty, a cross-sectional UQR model studies how changes in the share of mothers in the workforce would alter quantile values of women's overall wage distribution. Using such

estimates, we can infer whether wage changes due to motherhood increase wage inequality between women. However, this is an entirely different research question than what the motherhood penalty literature sets out to study. The UQR model examines how, for example, the median of all women's wages would change if the share of mothers relative to all women increases. In contrast, QTE models investigate the differences in median wages between mothers and childless women.

The origin of this mismatch builds upon a misunderstanding that UQR solves an issue of the traditional conditional quantile regression (CQR) model (Koenker and Bassett Jr, 1978; Koenker, 2005). The CQR model estimates *differences* between conditional quantile values. For example, the 90th percentile of the wage distribution is considerably higher for white women with an advanced college degree and ample work experience than young black women without high school. When controlling for education, race, and experience, the CQR model removes the influence of such variables and estimates motherhood effects on conditionally high or low quantiles. Since the estimates of a CQR model alone do not provide any information about the location of the group-specific percentile values within the overall distribution, its use for estimating quantile treatment effects is often limited (Melly and Wüthrich, 2017; but see Machado and Mata 2005; Melly 2005a). Concerning this point, the seminal paper of Killewald and Bearak (2014), and the later paper of Wenz (2019), was timely and vital in increasing the awareness of how CQR models should be interpreted.

However, the approach that replaced the CQR model in many research streams is ill-suited to investigate quantile treatment effects. Despite an early proposal for estimating the QTE model by Firpo (2007), the UQR model gained massive popularity across several disciplines because it *seemingly* tackled the CQR model's problem of *localizing* outcome differentials across the distribution (Killewald and Bearak, 2014; Wenz, 2019). However, the UQR model was developed to infer how changes in the independent variables influence unconditional quantile values (Firpo et al., 2009). The strength of an influence on the overall distribution is conditional on the independent variables' composition, the distribution of the outcome variable, and the unconditional quantile treatment effects. The UQR model's coefficients capture these factors' combined effect – called unconditional quantile partial effects (UQPE). Thus, the typical interpretation and use of UQR to estimate QTE within, for example, parts of the sociology literature (Budig and Hodges, 2014; Cooke, 2014; Killewald and Bearak, 2014; England et al., 2016; Glauber, 2018; Lin and Weiss, 2019; Wenz, 2019), the education literature (Porter, 2015), and the economics literature (Lindqvist

and Vestman, 2011; Havnes and Mogstad, 2015) is insufficient, and in worst case erroneous.

The purpose of this paper is to provide a gentle introduction to the differences between UQR and QTE models. As noted above, there is still confusion surrounding quantile regressions in the applied literature. Therefore, this paper fills a vital knowledge gap by highlighting what research questions these different models can answer. The first part of the paper addresses the conceptual difference between individual-level quantile treatment effects and influences on overall distributions, which we call population-level influences. Note that although we place both approaches in the context of causal analysis, the general ideas and insights apply equally to non-causally oriented analyses, such as purely descriptive studies.

The second part of the paper illustrates these conceptual differences between UQR and QTE models using a simple randomized experimental design setting (where control variables are not needed). The purpose is not to provide an exhaustive list of all scenarios in which quantile regression models differ. Instead, we pick two easy-to-communicate aspects that highlight the crucial differences between the models. Specifically, we show that UQR models respond differently than QTE models to variations of the treatment group's share and the shape of the overall outcome distribution (normal distribution vs. right-skewed distribution). The simulations convincingly demonstrate that the UQR model answers another type of question than the QTE models, even in a simple setting such as a randomized controlled trial, where there is no need to account for confounders.

The conceptual and empirical differences between various quantile regression models underline the need to match the correct quantile regression model to the specific research question. Thus, this paper adds to the growing awareness that failure to match research questions with the proper analytical strategies can lead to deeply misleading conclusions (Lundberg, Johnson and Stewart, 2021). Nevertheless, a key question from an empirical point of view is whether published studies using UQR have indeed drawn the wrong conclusions from their data. To check this, we compare quantile regression estimates on the motherhood wage penalty, family background effects, and scarring effects of unemployment.

Most importantly, we revisit the innovative motherhood wage penalty studies published in the American Sociological Review, which arguably sparked the use of UQR within sociological research (Budig and Hodges, 2010, 2014; Killewald and Bearak, 2014; England et al., 2016). Reassuringly, using two-way fixed effects models on a pooled panel data sample from the National Longitudinal Study of Youth (NLSY79), we

find that population-level influences of motherhood (estimated using UQR) are similar to the impact on individual-level wage penalties (estimated using QTE models). In contrast, differences are large in the two other examples. We end the paper by discussing why different quantile regression models' estimates diverge sometimes but not always, and conclude by providing some practical suggestions for future research.

Individual-Level Treatment Effects and Influences on the Overall Distribution

Individual-level: Average treatment effects and quantile treatment effects

Clarity about the research's target of inquiry, known as estimand, is paramount for the research contributing to knowledge accumulation, and the targeted estimand guides which estimation procedure is needed (Lundberg et al., 2021). A key difference between QTE and UQR models is that they are developed to estimate different estimands; the former concerns individual-level treatment effects, while the latter captures population-level effects. In this section, we introduce the concept of individual-level QTEs by contrasting it to the much more commonly used average treatment effect (ATE), whereas the next section contrasts these individual-level effects (QTE and ATE) with their population-level counterparts. We begin this section by focusing on the conceptual distinction between QTE and ATE based on examples with random treatment assignments before briefly commenting on how to estimate these quantities in more realistic settings.

Let us begin by defining the core concepts before illustrating these concepts using simulated data. In the potential outcomes framework (Morgan and Winship, 2007), the causal effect of a treatment for a single unit is defined as:

$$\delta_i = y_{i,1} - y_{i,0} \qquad [1]$$

with $y_{i,1}$ as the outcome for observation i given the treatment and $y_{i,0}$ for the same observation i without the treatment.

Using all observations within their two potential states, we could calculate the commonly used ATE as the difference between the average outcome given treatment and the average outcome (for the same individuals) given no treatment:

$$ATE = \sum_{i=1}^{n} \frac{\delta_i}{n} = \bar{y}_{i,1} - \bar{y}_{i,0} \qquad [2]$$

Causal effects are defined based on hypothetical states in a thought experiment, and observing both conditions for each individual at once is obviously impossible in reality (Holland, 1986). Nevertheless, although one of the outcomes is always missing, we can estimate

ATE by comparing the expected difference in the outcome between treated and non-treated under certain assumptions, such as unconfoundedness and the stable unit treatment value assumption (here T stands for treatment).

$$\widehat{ATE} = E[Y_1] - E[Y_0] = E[Y | T = 1] - E[Y | T = 0] \qquad [3]$$

Thus, the estimated ATE is calculated as the difference between the average of treated units' outcomes and the average outcome of non-treated ones. Notably, the assumptions needed depend on the chosen estimator; for example, when ATE of a binary treatment variable is estimated using OLS (rather than, for example, propensity score matching), average treatment effects on the treated and untreated must be identical to avoid bias (unless treatment and control groups are equally sized) (Słoczyński, 2022).

The QTE differs primarily from ATE in that it compares quantile values rather than means; otherwise, it is based on a similar line of thought as the ATE. If we know the whole distribution of the potential outcomes under the treated and untreated conditions, then we can compare quantile values in these potential outcome distributions. For example, we can compare the median of the outcome in the treated condition with the median in the non-treated condition to examine treatment strength at the median. (For now, we assume that both units would have the same rank within the counterfactual distribution in the absence of the treatment effect.) Thus, we can replace means, which are used to estimate ATE, with quantile values to calculate the QTE:

$$QTE_\tau = Q_{\tau,Y1} - Q_{\tau,Y0} \qquad [4]$$

where $Q_{\tau,Y1}$ and $Q_{\tau,Y0}$ are the values of the quantile τ for potential outcome distributions given the treatment condition and the no-treatment condition (Frölich and Melly, 2008; Melly and Wüthrich, 2017).

As noted above, since the counterfactual outcomes are unobserved, we need to compare treated and untreated units to estimate QTEs. Building upon the same generic assumptions as with ATE, we can estimate the QTE by comparing the realized outcome distribution of the treated with the corresponding outcome distribution of the non-treated:

$$\widehat{QTE_\tau} = Q_{\tau,Y1} - Q_{\tau,Y0} = (Q_\tau | T = 1) - (Q_\tau | T = 0) \qquad [5]$$

For example, we can calculate the QTE for the 95th percentile by comparing the value of the 95th percentile among treated individuals to the 95th percentile among the untreated, much the same as we calculate ATE by comparing the mean of the outcome in the treated and untreated group. In fact, without any control variables, this is what the CQR model estimates.[1]

Unlike ATE, QTEs can provide information about treatment effects for individuals located in different parts of the outcome distribution, given that an assumption called rank invariance, or the somewhat weaker assumption of rank similarity, holds (Dong and Shen, 2018; Frandsen and Lefgren, 2018). Rank invariance means that individuals maintain their expected ranks in the potential outcomes distributions. In that case, QTEs can be interpreted as treatment effects for individuals at quantile τ. For example, the estimated motherhood penalty at the 95th quantile identifies the motherhood impact for women that would have occupied the 95th quantile in the absence of motherhood.

The plausibility of the rank invariance assumption depends on the specific research question. However, although the rank invariance assumption enriches the interpretation of quantile regression results, it can often be relaxed. Specifically, the QTEs can be interpreted as comparing the same quantile in the treated and untreated distributions, rather than individuals occupying different positions within the potential outcomes distributions (Firpo, 2007; Melly and Wüthrich, 2017). Concerning the motherhood example, this would mean comparing the 95th percentile wages among mothers with the corresponding 95th percentile in the contrafactual childless scenario, where the same women may or may not occupy the 95th percentile in the two potential outcomes distributions.[2]

Estimating quantile treatment effects

Identifying QTEs is challenging in real-world settings without random assignment to the treatment or where we, for some other reasons, want to include control variables. Several scholars have noted that CQR is not well-suited to identify unconditional QTEs in the presence of control variables (Firpo, 2007; Killewald and Bearak, 2014; Porter, 2015; Wenz, 2019). There are, however, a few available approaches that identify unconditional QTEs adjusting for control variables.

To begin, Firpo (2007) provided an early two-step solution to estimating QTEs with a binary treatment variable and selection-on-observables, building on the propensity score matching framework. In the first step, a logistic or probit regression is used to estimate the likelihood of being treated. Then, a weighted CQR model is estimated in the second step, with weights identical to the inverse probability of being treated. Frölich and Melly (2010) extended Firpo (2007)'s approach, allowing for instrumental variables.

More recent approaches have provided solutions to non-binary treatment variables and fixed effects. Powell (2020) has developed the generalized quantile regression (GQR) model that allows for categorical and continuous variables and instrumental variables. Finally, Borgen et al. (2021b) have developed the residualized quantile regression (RQR) model, which in addition to handling all treatment variables, is computationally fast and can include high-dimensional fixed effects (Borgen, Haupt and Wiborg, 2021a). The RQR model solves the selection problem by regressing the treatment variable on the observed control variables, including fixed effects, using OLS. Then, the OLS residuals from the first step are used as the treatment variable in the second step CQR model.

It is beyond the aim of this paper to discuss these various estimation approaches in more detail. Here, our goal is to highlight the conceptual differences between different quantile regression models so that scholars can use the correct methodological tool for their research questions. Sociologists have a rich toolbox for estimating QTEs if that is their goal, with flexible, fast, and easy-to-implement models. Therefore, if the goal is to estimate QTEs, there is no reason to use the UQR model, which was not developed to estimate QTEs.

Population-level: Changes of the unconditional distribution

Individual-level treatment effects concern differences in outcomes between the same individual units given two treatment states. In contrast, population-level influences concern the consequences of the treatment on the outcome distribution, such as the overall inequality level (Bloome and Schrage, 2019). For example, we could ask how much the motherhood wage penalty contributes to the observed level of wage inequality. This question's answer depends partly on the share of mothers within a given population, as a sizeable penalty for a tiny group obviously results in a negligible influence on the overall distribution. However, in contrast to changes to the overall mean, which is straightforward to infer from individual-level treatment effects, the consequences on overall distributions are far more complex.

To illustrate the distinction between changes in population-level means and quantile values, we simulate two scenarios with random assignment to the treatment, where one-third of the individuals are treated. In both scenarios, the ATE is equal to four. However, while the treatment effect is constant across all individuals in the first scenario (Figure 1), it depends on the location of the treated within the outcome distribution (before the treatment) in the second one (Figure 2).

To begin with, let us examine how to translate the ATE into its population-level equivalence based on the simulated data in Figures 1 and 2. Panel A shows the potential outcomes under the treated and untreated conditions, and the average difference across the potential outcomes is four (i.e., $ATE = 4$). However, since only a third of the population receives the treatment
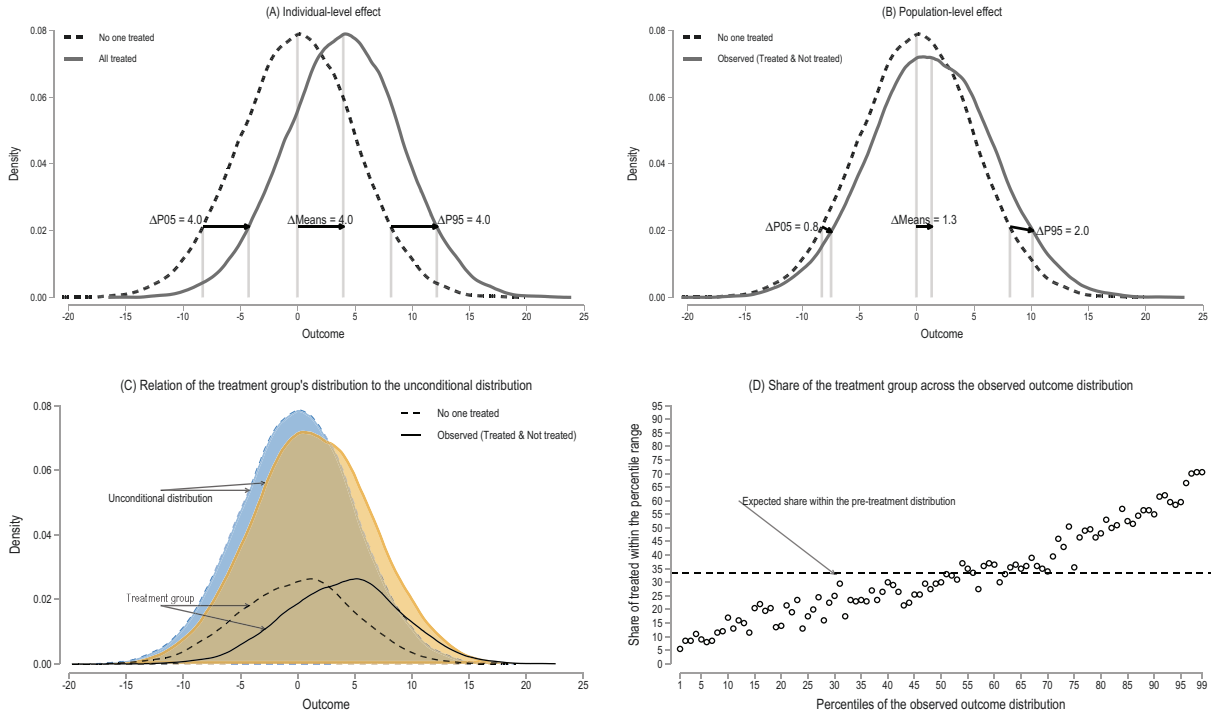
Figure 1 Simulated data to illustrate the distinction between individual-level effects and population-level effects in a scenario with uniform treatment effect. Note: The data in this graph consists of one draw of a sample of 20,000 individuals *i*. The simulated data is generated in the following manner: First, we draw a random pre-treatment outcome variable that is normally distributed with a mean of 0 and a standard deviation of 5 ($y_i^0 \sim N(0, 5)$). Second, we randomly assign the binary treatment T_i to 1/3 of the sample. Third, we define the post-treatment outcome variable as $y_i^1 = y_i^0 + \eta * T_i$, where the treatment strength η is 4 for all units. Note that we consider the case with full compliance and do not simulate stochastic deviations of treatment effects. 'No one treated' in panels A–C refers to a population where none have received the treatment. 'All treated' in panel A refers to a population where everyone is treated, whereas 'Observed (Treated & Not treated)' refers to a population where 1/3 are treated. The supplementary online appendix includes Stata do-files to reproduce the graph.

in this example, the treatment cannot increase the observed outcome distribution's overall average by four units. Instead, the change of the overall mean from the counterfactual state in which no one was treated to the observed state (with 1/3 treated) is equal to the changes of the means caused by treatment status, weighted by the probability of being a member of the treatment or control group.

$$\Delta \overline{Y} = (\overline{Y}_{i1} - \overline{Y}_{i0}) * P(T = 1) + (\overline{Y}_{i0} - \overline{Y}_{i0}) * P(T = 0) \quad [6]$$

In both of our simulation scenarios, the treatment shifted the overall mean of the outcome distribution upwards by 1.333 since the treatment affected a third of the units and the average treatment strength was equal to four for all units.[3]

Unfortunately, the same simple logic cannot be applied to translate individual-level treatment effects into population-level effects on quantile values. Let us

show this fact empirically before providing an intuition of why influences on quantile values are more complicated than influences on means. Panel B of Figures 1 and 2 show the counterfactual distribution without treatment (dashed line) and the observed distribution, consisting of a mix of treated and untreated individuals (solid line). Moreover, it displays changes in the mean, 5th percentile, and 95th percentile of the overall outcome distribution. Even with uniform treatment effects across all units (Figure 1), the treatment increases the 5th and 95th percentile values by 0.8 and 2.0, respectively. Importantly, these different population-level effects on quantile values appear despite the same quantile treatment effects for all and a random treatment assignment.

There are two important takeaway points from panels A and B of Figure 1. First, the treatment's effect on the individual level differs markedly from its consequences on the population level, even in a

Scenario 2: Increasing treatment effect across the distribtion
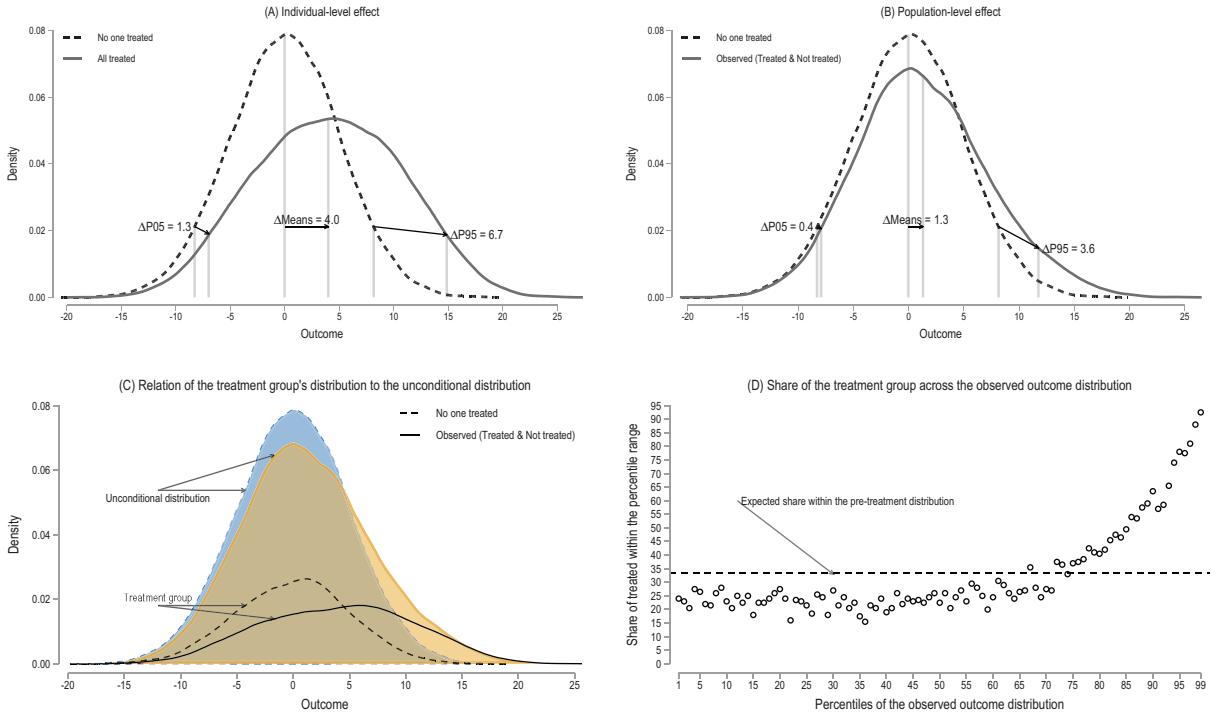
Figure 2 Simulated data to illustrate the distinction between individual-level effects and population-level effects in a scenario with increasing treatment effects across the distribution. Note: The simulated data in this graph is generated similarly as in scenario 1 (see table note in Figure 1), except the post-treatment outcome variable is defined as $y_i^1 = y_i^0 + \eta * T_i$, where $\eta = 1 + 6 * r_i$ and r_i defines the rank of the individual i in the pre-treatment outcome distribution y_i^0 and follows a discrete uniform distribution that ranges from 0 to 1 ($r_i \sim U[0, 1]$). As an illustration, the treatment strength (η) is equal to $1 + 6 * .05 = 1.3$ for units located at the 5th percentile of the pre-treatment outcome distribution and $1 + 6 * .95 = 6.7$ for units located at its 95th percentile. The average treatment strength summed over all quantiles is 4.0, as in scenario 1. The supplementary online appendix includes Stata do-files to reproduce the graph.

simplistic scenario. Second, since the treatment effect is constrained to be the same for all units, the treatment's increasing influence on the unconditional quantile value is not based on a rising quantile treatment effect. This conclusion warrants the question of why a uniform treatment effect translates into heterogeneous impacts on the unconditional outcome distribution.

Compared to the counterfactual state where none is treated, treating a third of the population changes many of the individuals' ranks throughout the observed outcome distribution. (Note that this is not the same as rank invariance, which is ranks in potential outcome distributions.) Let us revisit the potential outcomes framework and consider the case of the 95th percentile value. A third of all units with an outcome equal to or larger than the 95th percentile in the counterfactual state where none is treated will later receive the treatment because of the random assignment (i.e., $(Y_1, Y_0) \perp T$). However, once a third

of the population is treated, giving us the observed outcome distribution, the share of treated units at the 95th percentile or above changes. As illustrated in panel D of Figure 1, the share of treated units between the 95th and 96th percentile is about 75% in the observed distribution (compared to 33% to-be-treated in the counterfactual one). Thus, the treatment increased the treated's outcomes and, in the process, changed their rank from the pre-treatment counterfactual distribution to the observed distribution. If the treatment pushes treated into the upper 5%, some untreated must leave the upper 5% – otherwise, the 95th percentile would no longer be a separator for the upper 5%. The newly entered treated units typically have higher outcomes than the upper 5% of the counterfactual distribution without treatment. Thus, the treatment increased the overall bar to be part of the top 5% of the observed outcome distribution.

The case for the fifth percentile value is even more complex. The treatment increased individuals'

outcomes with low pre-treatment outcomes by four units, placing them towards the middle of the observed distribution and reducing the share of treated units within the lower 5% of the observed distribution sharply (Panel D of Figure 1). In the upper 5% case, the treatment increased outcomes, thereby reducing the likelihood of the non-treated being part of the upper 5%. In contrast, the treatment pushed almost all treated units out of the lower 5%.

Still, the outcomes of the non-treated did not change, which are now the majority within the lower 5% (and the majority in the population). The value separating the lowest 5% of all units increases slightly, as the lower 5% of the observed distribution includes some non-treated with higher outcomes; nevertheless, it grows less than one would expect given the quantile treatment effect only. The low impact of the treatment on the lower end of the outcome distribution is because relatively few units are treated overall (33%). Thus, the movement of some treated out of the lower 5% does not influence the overall distribution as much as the massive inflow of treated into the upper rungs. This comparison highlights the importance of the treated group's size for the influence on the overall distribution. At this point, it is worth mentioning that the composition effect disappears when *all* units are treated, in which case the population-level influence on quantile values coincides perfectly with the QTE. This similarity is because the underlying individual-level effect does not lead to a resorting of ranks.

Above, we discussed the case of a constant treatment effect; we now briefly consider the case of an increasing quantile treatment effect (Figure 2). On the population level, increasing QTEs changes the share of the treated across the distribution differently compared to the uniform treatment effect (contrast panel D in Figures 1 and 2). In the second scenario, we observe more treated units in the lower and upper parts of the distribution. Thus, the second scenario's increasing QTEs influence the lower parts of the outcome distribution less and the upper parts more than the first scenario's uniform QTEs.

The paragraphs above highlight that we cannot calculate the population-level changes in the overall quantile values based on the QTE and treated share. Scholars need to take into account that the treatment influences parts of the distribution differently – even if the treatment has a uniform impact on the individual level. Thus, to model the consequence of the treatment on the overall distribution, we need a method that captures the treatment's entire *influence* on the population level. This influence depends on the treatment effect, but it also depends on the distribution of the treated and the unconditional distribution's density at particular quantiles of interest.

Estimating influences on unconditional quantile values

Firpo et al. (2009) introduced the UQR model as a solution to estimate influences on unconditional distributions, which they call the unconditional quantile partial effect (UQPE). They proposed a two-step approach where scholars (1) re-center the influence function (RIF) and (2) regress the RIF on the independent variables in an OLS model.[4] The RIF is defined as

$$RIF(y; Q_\tau) = Q_\tau + \frac{\tau - \mathbb{I}\ (y \leq Q_\tau)}{f_Y(Q_\tau)} \qquad [7]$$

with $\mathbb{I}$ describing the indicator function,[5] $f_Y(Q_\tau)$ referring to the density of the outcome at quantile τ, and Q_τ to the sample quantile value of τ. The RIF has only two values, one for units with an outcome value below or equal to the quantile value Q_τ and one for units with an outcome value large than Q_τ. Regressing the RIF at a specific quantile on the independent variables gives the UQPEs, further detailed in Appendix A.

$$\widehat{\beta} = \mathrm{E}\left[RIF(y; Q_\tau)\ |\ T = 1\right] - \mathrm{E}\left[RIF(y; Q_\tau)\ |\ T = 0\right] \qquad [8]$$

From a technical point of view, UQR compares two counterfactual distributions: one where all units are treated and one where none are treated. Thus, the UQR coefficient refers to changes from 0 to 100% treated. However, most scholars estimating UQPEs using the UQR model are not interested in such counterfactuals. Instead, they want to know how some groups contributed to the overall outcome distribution, such as overall wage inequality. The treatment's contribution to the observed distribution can partly be recovered by multiplying the UQR coefficients by the proportion treated (Fortin, Lemieux and Firpo, 2011; Rios-Avila, 2020). That said, the UQR model may produce biased estimates of the UQPE when used to study considerable changes in the distribution of the independent variables, which is particularly relevant for binary predictors (Firpo et al., 2009; Rothe, 2009, 2012; Firpo and Pinto, 2016; Rios-Avila, 2020).

The various sources of potential bias in the estimated UQPEs and QTEs, including confounding, are outside the scope of this paper; our main aim is to highlight the crucial difference between individual-level QTEs and population-level UQPEs. Thus, in the following data simulations, we ignore potential bias in the UQPE and emphasize the differences in raw coefficients.

Illustrating the Difference Between the UQR Model and QTE Models Using Simulated Data

Data generating process

This section illustrates the differences between the individual-level quantile treatment effects estimated by

QTE models and the population-level influences estimated by the UQR model using further data simulations. In the data simulations, we vary the distribution of the treatment variable and the outcome variable's distribution, which are two key parameters that may cause UQPE to differ from QTE. Applications of UQR differ in the size of the treatment group (migrants, gender, policy changes for all) or the distribution of interest (bell-shaped, like income or test scores, or skewed, like wealth or health measures). We suspect that erroneous conclusions in studies that have used UQR to study QTE can likely be traced back to treatment group size or distributional shape.

In the main simulations, all treatment variables are binary (d_i^p). We vary the probability p of the dummy treatment variable d^p being equal to 1 from 90 to 10%.[6] The data simulation consists of running 1,000 draws of $N = 1,000$ for four simulation scenarios.

In scenario 1, we generate data where the QTE does not vary over the outcome distribution (i.e., d_i^p produces a location shift only) and where, accordingly, the QTEs are identical to the average treatment effects (ATE) from OLS for all quantiles ($QTE^\tau = OLS \; \forall \; \tau$). First, we draw a random pre-treatment outcome variable that is normally distributed with a mean of 0 and a standard deviation of 1 ($y_i^0 \sim N(0, 1)$). Second, we assign the binary treatment variable d_i^p to p percent of the sample. Third, we define the post-treatment outcome variable y_i, which will serve as the outcome in the simulation analyses below, as:[7]

$$y_i = y_i^0 + \beta * d_i^p; \quad where \; y_i^0 \sim N(0,1) \; and \; \beta = 1 \tag{9}$$

In scenario 2, we generate data where the QTEs vary over the outcome distribution. We do this by allowing the strength of the treatment variable (d_i^p) to depend on the individual i's percentile rank (r_i) in the pre-treatment outcome distribution (y_i^0).[8]

$$y_i = y_i^0 + (2 * \beta * r_i) d_i^p; where \; y^0 \sim N(0,1)$$
$$and \; r_i \sim U[0,1] \tag{10}$$

The setup in scenarios 3 and 4 is the same as in scenarios 1 and 2, respectively, except the outcome is a right-skewed variable. In scenarios 3 and 4, we replace $y_i^0 \sim N(0,1)$ from equations 9 and 10 with $y_i^0 \sim Gamma(1,2)$. The outcome variable's distributional shape influence the density of the distribution at various quantiles. The density matters because the treatment's influence on an overall quantile value increases the lower the density around the quantile is. Suppose we have a skewed distribution such as wealth, where large parts have low density, and others have sizeable densities. In that case, a uniform QTE would likely translate into a differential UQPE across the outcome distribution.

We compare coefficients using the UQR model with coefficients using the CQR model in all four scenarios. Note that while methods to identify unconditional QTEs in the presence of control variables have been developed (Firpo, 2007; Frölich and Melly, 2010; Powell, 2020; Borgen et al., 2021b), these methods trivially provide the same point estimates as the CQR model without any control variables. However, as the simulations demonstrate, the same is not the case for the UQR model since this model identifies influences on overall quantile values (i.e., UQPE) rather than individual-level QTEs.

The backdrop of this paper is the widespread use of the UQR model to estimate QTEs. Given this practice, the main objective of the simulations is to highlight that individual-level quantile treatment effects estimated by QTE models differ from the population-level influences estimated by the UQR model. These coefficient differences reflect the fact that population-level influences differ from individual-level QTEs.

We perform all simulations in Stata 16.1 and use the rifreg command to estimate UQR coefficients (i.e., RIF-OLS) and the qreg command to estimate QTE coefficients. Supplementary Online Appendix D includes Stata codes to reproduce the results.

Share of treated individuals and the distribution of the outcome

Figure 3 presents the main simulation results from scenarios 1 to 4 with the QTE coefficients in the top row and the UQR coefficients in the bottom row. Under the faulty assumption that the UQR model identifies the same estimand as a QTE model, the top row results should be identical to the bottom row results. However, a quick inspection of Figure 3 shows that this clearly is not the case.

In simulation scenario 1, the QTE is equal to 1 at all quantiles, as shown by the QTE coefficients (panel A). Contrary to this, a constant QTE leads to a highly asymmetric influence on overall quantile values (i.e., UQPE) across the distribution, as estimated by UQR (panel E). The UQR coefficients vary across the outcome distribution, with the trends being diametrically opposite depending on the treatment group's size; the UQR coefficients decrease across the outcome distribution when 10% is treated and increase when 90% is treated. Thus, we cannot infer from the structure of the UQR results on QTEs, which showcases the risk of misinterpreting the results when using UQR to study QTEs. Importantly, this is the case even in the simplistic scenario with a random assignment to the treatment variable, a normally distributed outcome variable, and a treatment effect that is restricted to be identical for all.

In scenario 2, we simulate data where QTE increases monotonically across the outcome distribution. Again,
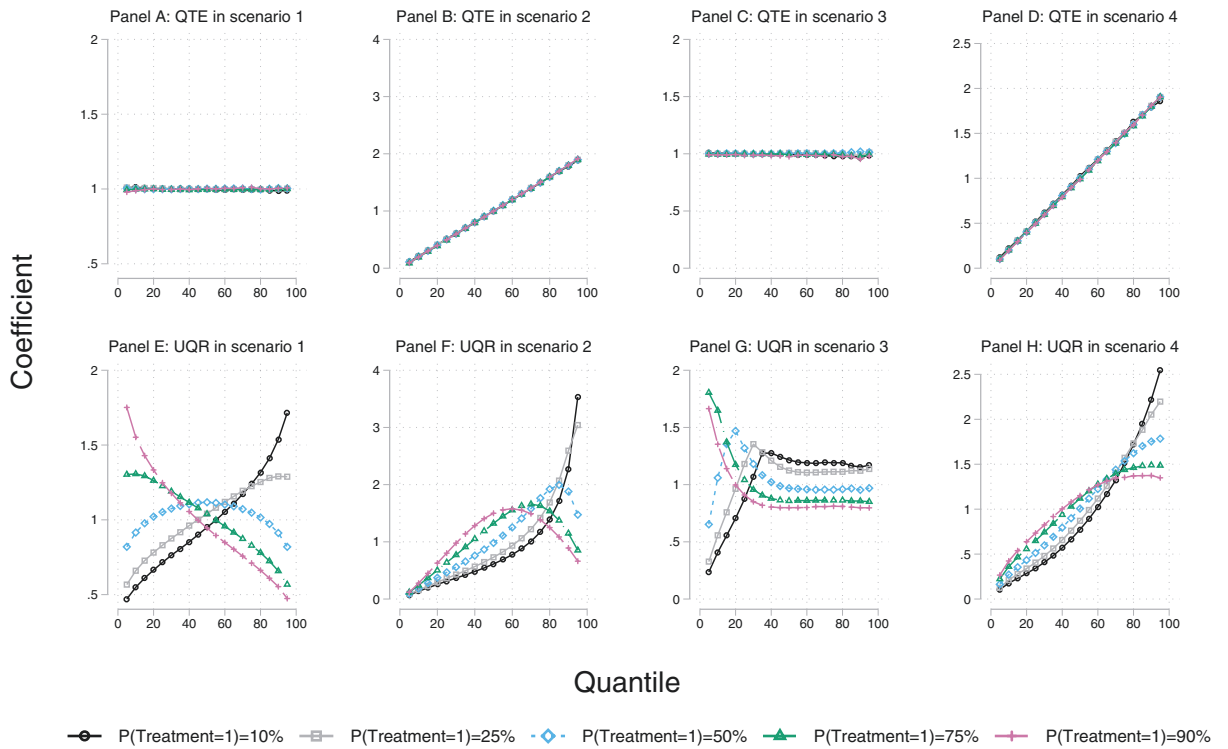
Figure 3 Comparing the QTE model (Panels A-D) and UQR model (Panels E-H) from 1,000 draws of *N* = 1,000 from scenarios 1–4 (scenarios 1 & 2: normally distributed outcome; scenarios 3 & 4: skewed outcome). Each line represents a dummy variable with 10, 25, 50, 75, or 90% having the value 1, and is estimated in separate models. Note: See Online Appendix Table B1 for detailed coefficients, including the standard deviation of the coefficients from the random draws, and Appendix Figures B1 and B2 for the distribution of coefficients for selected quantiles.

the estimated UQPE (panel F) differs considerably from the estimated QTE (panel B). When 10 or 25% is treated, UQR coefficients increase across the outcome distribution but in a much more dramatic fashion than for QTE. Interestingly, comparing the UQR coefficients in scenarios 1 and 2 highlights that we cannot use the UQR model to inform whether the QTEs are uniform (scenario 1) or increasing (scenario 2) when the share of treated is few. Turning to the case where 75% or 90% is treated, the UQR coefficients show an inverted U-shape. In sum, different treatment composition leads to substantially different results.

In scenario 3, where the outcome is right-skewed, the QTE is again restricted to be the same across all quantiles. The estimated UQPEs (panel G) differ from the estimated QTEs (panel C), especially in the bottom third of the outcome distribution, and the share of the treated individuals affects the estimated UQPE. The UQR estimates in panels G and E also differ fundamentally – despite the same underlying QTE. The differences highlight the importance of the density for the treated's influence on unconditional quantile values, which is an important part of the construction of

the RIFs. This result should further caution against an inference from UQR results to QTEs. The same QTEs may translate into vastly different UQPEs depending on the outcome variable's distributional shape. Finally, in scenario 4, the QTEs increase monotonically across the outcome distribution (panel D) while the UQPE differs somewhat depending on the size of the treated group (panel H), although less than in the other three scenarios.

Treatment strength

The simulation results above illustrate that UQR estimates may differ from QTE estimates. However, the differences are not always as striking. Recall that a key reason why the UQR model differs from the QTE models is that the ranking of the individuals differs in the counterfactual distributions; for example, the treatment pushes some treated individuals out of the bottom 5% of the outcome distribution and raises the bar to be among the top 5%. In this respect, the treatment strength matters because changes in the ranking are less likely if the treatment effects are small.

To illustrate this point, we have re-run simulation scenario 1 (described above), holding treatment effects fixed but varying the proportion of the variance in the outcome explained by the treatment (by varying the residual variation in the outcome) (Figure 4). This exercise shows that the differences between UQR and QTE coefficients increase when the relative strength of the treatment increases, with differences in this specific data simulation being mostly negligible when the treatment explains less than 1% of the variance in the outcome.
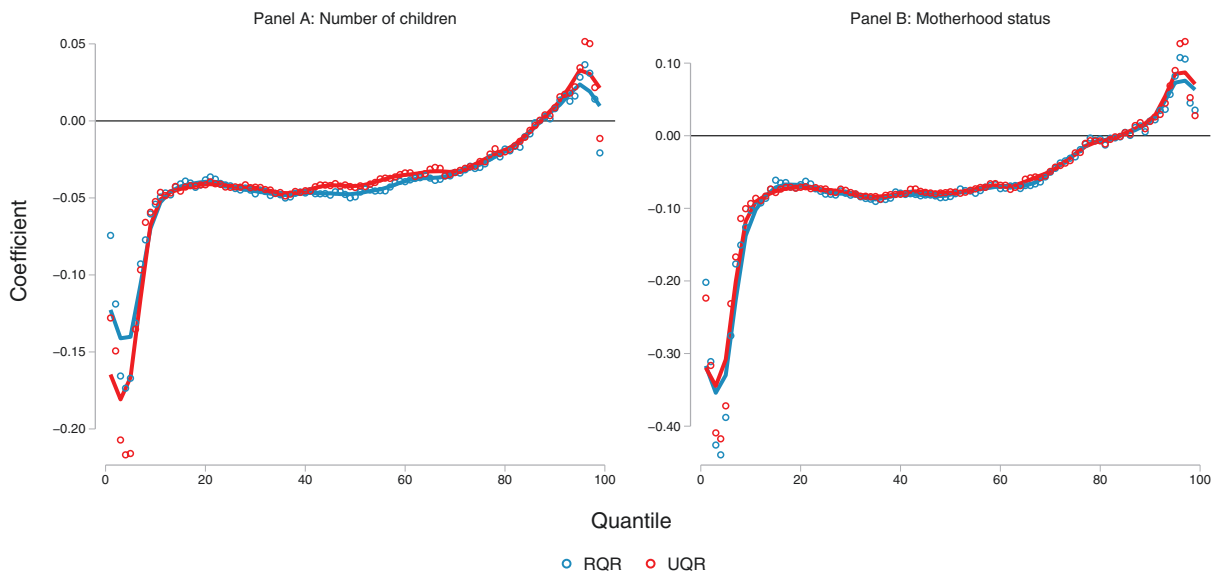
The Motherhood Wage Penalty Revisited

The oft-overlooked distinctions between quantile regression models are not merely technical, trivial differences. Failure to match the correct quantile regression approach to the specific research question may, in some cases, result in wrong conclusions being drawn from the data and, in the end, misguided theories. The simulations above showed that the conceptual and methodological differences between quantile regression models could manifest themselves in diverging empirical patterns. This section revisits the motherhood wage penalty to examine whether the frequent use of the UQR model in this literature has led to wrong conclusions.

The motherhood wage penalty example is not chosen at random. The novel motherhood wage penalty studies sparked the use of UQR within broader sociological research (Budig and Hodges, 2010, 2014; Killewald & Bearak, 2014; England et al., 2016). These studies used two-way fixed effects models on pooled panel data from the National Longitudinal Study of Youth (NLSY79). It started with Budig and Hodges (2010), who used CQR to show that the motherhood wage penalty is larger at the bottom of the earnings distribution than at the top. Killewald and Bearak (2014) later (correctly) criticized their use of CQR and (incorrectly) proposed to use the UQR model to estimate how the motherhood gap varies across the unconditional wage distribution. They found the largest penalties at the middle of the distribution and the smallest at the top, while Budig and Hodges (2014), in their reanalysis using the UQR model, found the largest penalties in the bottom half.
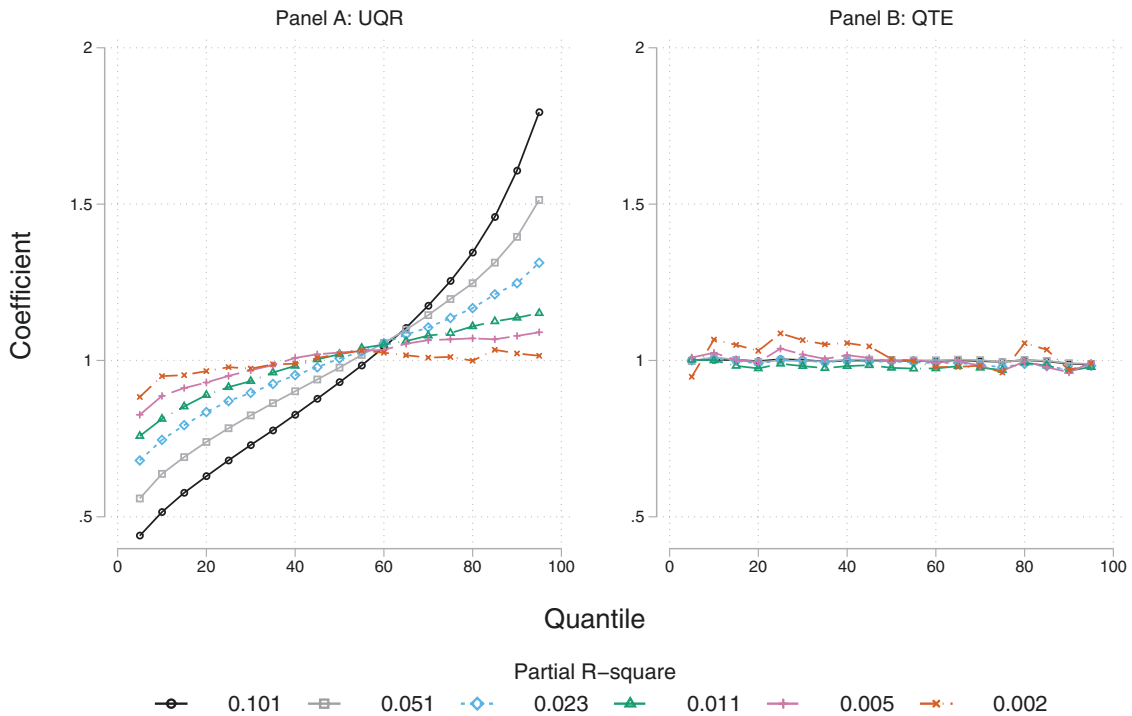
Figure 4 Comparing UQR (panel A) and QTE (panels B) coefficients of a binary treatment variable (10% = 1) in simulation scenarios with varying treatment effects in terms of the treatment's partial R^2 (2,000 draws of $N = 1,000$). Each line represents coefficients in simulation scenarios with a given treatment strength, expressed as the proportion of the outcome's variance explained by the treatment. Note: The simulation setup is similar to simulation scenario 1 (see equation 9), except that the residual variance across the simulations is set to 0.90^2 ($R^2 = 0.10$), 1.3^2 ($R^2 = 0.051$), 2^2 ($R^2 = 0.023$), 3^2 ($R^2 = 0.011$), 5^2 ($R^2 = 0.005$), and 9^2 ($R^2 = 0.002$). In the main text, the residual variance was 1, which translated into a partial R^2 of 0.082.

Figure 5 The estimated size of the motherhood wage penalty using an unconditional QTE model (RQR) and the UQR model in a sample of non-Hispanic white women from the NLSY79 (*N* = 3313 persons with 43.768 person-years). Note: Scatter points represent the estimated UQR (red dots) and RQR (blue dots) coefficients for quantiles 1–99 in steps of 1. The red and blue lines plot a local polynomial smooth of the coefficient on quantile. All models are estimated with controls for education, experience, tenure, usual work hours, full-time worker, age, enrolled in school, marital status, spouse's earnings, spouse's work hours per week, region, and rural region. The UQR coefficients are estimated using the rifhdreg command (Rios-Avila, 2020), and the RQR coefficients are estimated using the rqr command (Borgen et al., 2021a). Supplementary Appendix D includes information to reproduce our results.

We contribute to the motherhood wage penalty discussion by matching the estimand with a correct quantile regression estimator. As noted in the introductory section, the motherhood wage penalty is best answered based on individual-level QTEs. Scholars want to know whether the wage gap between mothers and childless women differs across the wage distribution. The UQR model informs on population-level influences on the unconditional wage distribution rather than such individual-level wage gaps. Furthermore, at the time of the ASR publications, methods that estimated QTEs in the presence of high-dimensional fixed effects did not exist. Since then, quantile extensions of classical fixed effects models have been developed (Borgen et al., 2021a, 2021b). These developments allow us to account for unobservable time-constant individual traits within a QTE framework, which has yet to be done in the motherhood wage penalty literature.

Using the pooled NLSY79 panel data provided by England et al. (2016), we estimate quantile regressions with individual-level fixed effects on a sample of non-Hispanic white women in the 1979–2010 waves (see note in Figure 5 for details). Our model specification resembles that of England et al. (2016) but is more parsimonious, and the results are not meant as a strict replication. Moreover, we focus solely on the choice of quantile regression estimator and ignore several other

unresolved challenges. First, our specification includes post-treatment variables that could be affected by the treatment, which may bias the penalties due to over-control and collider bias (Elwert and Winship, 2014). Furthermore, our specification ignores the dynamics over time in wage growth (Ludwig and Brüderl, 2018), and it ignores potential challenges that could arise by pooling data over many years.[9] Future empirical applications should tackle these challenges using an appropriate quantile regression model to identify unbiased motherhood QTEs.

Reassuringly, unlike the simulation results above, the motherhood QTE and UQR coefficients follow the same pattern across the wage distribution (Figure 5). Using both models, we find that the effects of number of children (continuous predictor) and motherhood status (binary predictor) are large for low-wage women but close to zero for high-earners (except in the very top, where the penalty again increases). For a woman with a wage equal to the 5th percentile, becoming a mother reduces her wages by about 15% for each additional child. The corresponding figure for a woman at the median or higher is less than 5%. These patterns across the distribution are consistent with Budig and Hodges (2010)'s initial results (see also Budig and Hodges, 2014; Killewald and Bearak, 2014) but are not compatible with the claim that high-skilled, high-income women face the largest motherhood penalty.

The striking similarity of the motherhood UQR and QTE coefficients raises the question of why these models differ sometimes but not always. Although several factors may come into play, one key reason is the relative treatment strength. The sizeable motherhood wage gap is important from theoretical and policy perspectives; yet, motherhood still explains only a small fraction of the variation in wages among white women. As illustrated in the simulations above, the differences between UQR and QTE estimates are negligible when the relative treatment strength is small.

However, the motherhood wage penalty example does not imply that one should infer QTEs from the UQR model. Overall, the simulation results illustrate that UQR estimates differ from QTE depending on the treated group's size and the distribution of the outcome variable. The Online Appendix C supplements this story by showing two empirical examples where the proportion treated and the outcome distribution's shape impacts the differences between UQR and QTE coefficients. While the motherhood case was chosen because much of the sociological conversation around quantile regression has been spurred by this field, the two supplementary cases are chosen because they highlight how UQR and QTE coefficients may differ in real-world applications. The first case concerns socioeconomic differences in academic achievements in the Panel Study of Income Dynamics data (Appendix Figure C1). UQR and QTE models give opposite patterns across the achievement distribution when comparing children with high socioeconomic backgrounds with other children (similar to panels A-B and E-F in Figure 3). The second concerns the scarring effects of unemployment on tenure in General Social Survey data, where tenure is a heavily right-skewed variable (Appendix Figure C2). At the top of the distribution, the estimated unemployment coefficients using UQR differ markedly from those using QTE.

Discussion

Although several studies have emphasized the typical misinterpretation of the CQR model in the last decade (Killewald and Bearak, 2014; Porter, 2015; Wenz, 2019), the crucial distinction between the UQR model and QTE models has received little attention across a range of disciplines. Consequently, there is a mismatch between the quantile regression models used in many studies (the UQR model) and these studies' aims (identify unconditional QTEs). Failure to link the relevant estimand to the correct estimation strategy could lead to misleading conclusions and eventually misguided theories (Lundberg, Johnson and Stewart, 2020). This paper has clarified the conceptual distinction between the individual-level quantile treatment effect identified

by QTE models (Firpo, 2007; Powell, 2020; Borgen et al., 2021b) and population-level influences on quantile value identified by the UQR model (Firpo et al., 2009). This concluding section discusses the implications for studies that have erroneously used the UQR model and provides guidelines for future work.

Although the UQR model was not designed to identify individual-level quantile treatment effects, a key question from an empirical point of view is whether published studies using UQR have indeed drawn the wrong conclusions from their data. Our data simulations illustrate the importance of matching the specific research question to the correct quantile regression approach, as different classes of quantile regression models may give entirely different results (to different questions). Nevertheless, the absolute difference in the estimated coefficients from these approaches may, to some extent, depend on the relative strength of the treatment on the outcome variable. The differences between QTE and UQR coefficients diminish as the individual-level treatment effects shrink toward zero.

That said, we firmly believe that future research should match the research question to the correct quantile regression toolkit. Admittedly, the chief aim from an applied perspective is to provide the correct answer to some research question, and to that end, some 'methodological pragmatism' may be a strength. However, for researchers interested in QTEs, there is no need to take the risk of biased estimates, as methods exist that allow for the inclusion of covariates without changing the coefficients' interpretation (discussed below). Further, this study has not provided an exhaustive list of all scenarios, and predicting the implications in complicated model specifications is difficult.

For that reason, we strongly recommend replicating key studies that have used UQR to estimate unconditional QTEs to check whether the results are sensitive to the choice of quantile regression estimator. Our reanalyses of the motherhood wage penalty in the NLSY79 data showed that QTE and UQR models lead to similar conclusions concerning how low- and high-earners are affected by motherhood. However, socioeconomic achievement gradients and scarring effects of unemployment are examples where we show that mapping the correct method to the research question is essential (see Online Appendix C).

It may be tempting to read this paper as a cautionary tale about how the decade-long spread of UQR within sociology and related fields to study individual-level treatment effects showcases the potential perils of easy access to advanced regression techniques. However, such reading undervalues the powerful ripple effects of novel methods use. The initial quantile regression paper in the American Sociological Review (Budig and Hodges, 2010) and the subsequent quantile

regression debate (Budig and Hodges, 2014; Killewald and Bearak, 2014; England et al., 2016) stimulated other researchers to go beyond averages to get complete views of associations between variables. The quantile regression approach's surging popularity contributed to substantial attention being directed towards further developing these methods. In the end, the reward is better empirical predictions and ways of testing theories against data.

In conclusion, we offer some general practical guidelines that help researchers match the research question to one of the three overall classes of quantile regression estimators. (For an overview of Stata packages to estimate the models, see Borgen et al., 2021a). First, research questions related to the payoff to characteristics across the unconditional outcome distribution fall under the category of unconditional QTEs and should be studied using QTE models. Besides the parenthood wage penalty example discussed above, other examples of such research questions are whether the benefit of payroll taxation rules across the income distribution (Haupt and Nollmann, 2021), wage returns to education (Borgen, 2015; Balestra and Backes-Gellner, 2017), and socioeconomic achievement gradients (Grätz and Wiborg, 2020).

The estimation of such unconditional QTEs depends on the model specification. In models without any control variables, different QTE models trivially produce the same results, and one could use the standard CQR algorithm (Koenker, 2005). In the presence of control variables, unconditional QTEs can be estimated using the propensity score framework introduced by Firpo (2007) if the treatment is binary (Frölich and Melly, 2010). In contrast, the generalized quantile regression (GQR) model of Powell (2020) and the residualized quantile regression (RQR) model of Borgen et al. (2021b) can be used with all types of treatment variables (Borgen et al., 2021a). Note that including high-dimensional fixed effects may be problematic in a propensity score framework. In such cases, the RQR model provides a straightforward extension of classical linear fixed effects models, whereas one could use the GQR panel estimator of Powell (2022) to include non-additive fixed effects.

Second, research questions examining influences on unconditional distributions fall under the unconditional quantile partial effects category and could be analyzed using the UQR model. A few examples are analysis of factors influencing income inequality (Biewen and Seckler, 2019), comparisons of student achievement distributions (Loyalka et al., 2019), or studies into how countries differ regarding the gendered division of housework time (Moreno-Colom, 2017). The UQR model could be used to examine influences on distributions irrespective of the types of included control variables (Firpo et al., 2009; Borgen, 2016; Rios-Avila, 2020).

Finally, some research questions aim at examining differences in distributions between groups, such as whether wages of public-sector workers are more homogenously distributed compared to private-sector ones (Melly, 2005b) or whether we can explain differences in regional driving fatality distributions with differences in drinking and driving policies (Ying, Wu and Chang, 2013). Studies interested in such between-group differences in quantile values should employ the traditional CQR model (Koenker and Bassett Jr, 1978; Koenker and Hallock, 2001).

Notes

1 Comparable to OLS, we can also use quantile regressions for more descriptive purposes, with or without control variables. For example, with gender as the independent variable, quantile regression coefficients give the differences in quantile values (e.g., the 90th percentile) between men and women.

2 Notably, in contrast to CQR, a QTE model with a relaxed rank invariance assumption would still localize the treatment effect within the overall distribution.

3 $\Delta \overline{Y} = 4.0 * \frac{1}{3} + + 0.00 * \frac{2}{3} = 1.333.$

4 The seminal paper of Firpo et al. (2009) introduces three ways to estimate UQR models: RIF-OLS, RIF-logit, and RIF-NP. These three different estimators often produce similar results (Firpo et al., 2009, p. 966), and the RIF-OLS model is preferred because of its ease of estimation and computational efficiency. The RIF-OLS approach builds upon the common 'folk wisdom' that a linear probability model (LPM) yields very similar estimates of average marginal effects (AME) as logit and probit models, and therefore does not need to estimate AME directly (Firpo, Fortin and Lemieux, 2007). Note also that the kernel density estimate depends on the bandwidth and kernel, which introduces some uncertainty that should be accounted for in the estimation of standard errors. These issues are not essential to our argument and will not be discussed here.

5 An indicator function is a function that checks whether a statement is true under a given condition and returns the value 1 in this case and 0 in all other cases. It is thus just a statement about the construction of a dummy variable.

6 In the Online Appendix Table B1, we also present simulation results with a uniformly distributed continuous treatment variable ($\sim U(0,1)$), a categorical treatment variable (4 equally sized categories), and a skewed treatment variable ($\sim Beta(2,25)$). Using a uniformly distributed continuous variable or categorical variable leads to similar results as with the evenly split binary variable (d_i^{50}), while the skewed variable does not.

7 Across all simulation scenarios, the model assumes homoscedasticity ($Var(y_i^0|d_i^p) = \sigma^2$). Further, the increasing QTEs in scenarios 2 and 4 does not increase the variance of the error term (y_i^0) compared to scenarios 1 and 3, respectively.

8   We multiply β wiith 2 in scenario 2 to make sure that the average treatment effect is equal to 1 in both scenario 1 and scenario 2.

9   Some applications of UQR use pooled samples (as in our example). One potential pitfall is that in UQR, the dependent variable is constructed using quantiles and densities from the pooled sample. If the outcome distributions change strongly over time, for example due to recessions, the quantile values and densities of the time-specific and the pooled sample do not align. An UQR model then captures the influence of covariates, like year dummies, on this overall, pooled distribution, in which time-specific distributions are located at different points.

Supplementary Data

Supplementary data are available at *ESR* online.

Acknowledgements

Funding

Data availability

The data underlying this article are available in the article and in its online supplementary material. See Appendix D for more details.

References

Balestra, S. and Backes-Gellner, U. (2017). Heterogeneous returns to education over the wage distribution: Who profits the most? *Labour Economics*, **44**, 89–105.

Biewen, M. and Seckler, M. (2019). Unions, internationalization, tasks, firms, and worker characteristics: A detailed decomposition analysis of rising wage inequality in Germany. *The Journal of Economic Inequality*, **17**, 461–498.

Bloome, D. and Schrage, D. (2019). Covariance regression models for studying treatment effect heterogeneity across one or more outcomes: understanding how treatments shape inequality. *Sociological Methods & Research*, **50**, 1034–1072.

Böheim, R. and Taylor, M. P. (2002). The search for success: do the unemployed find stable employment? *Labour Economics*, **9**, 717–735.

Borgen, N. T. (2015). College quality and the positive selection hypothesis: The 'second filter'on family background in high-paid jobs. *Research in Social Stratification and Mobility*, **39**, 32–47.

Borgen, N. T. (2016). Fixed effects in unconditional quantile regression. *Stata Journal*, **16**, 403–415.

Borgen, N. T., Haupt, A., and Wiborg, Ø. (2021a). Flexible and fast estimation of quantile treatment effects: the rqr and rqrplot commands. *SocArXiv*, doi: 10.31235/osf.io/4vquh. preprint: not peer reviewed.

Borgen, N. T., Haupt, A. and Wiborg, O. (2021b). A new framework for estimation of unconditional quantile treatment effects: The Residualized Quantile Regression (RQR) model. *SocArXiv*, doi: 10.31235/osf.io/42gcb. preprint: not peer reviewed.

Budig, M. J. and Hodges, M. J. (2010). Differences in disadvantage: variation in the motherhood penalty across white women's earnings distribution. *American Sociological Review*, **75**, 705–728.

Budig, M. J. and Hodges, M. J. (2014). Statistical models and empirical evidence for differences in the motherhood penalty across the earnings distribution. *American Sociological Review*, **79**, 358–364.

Cooke, L. P. (2014). Gendered parenthood penalties and premiums across the earnings distribution in Australia, the United Kingdom, and the United States. *European Sociological Review*, **30**, 360–372. doi: 10.1093/esr/jcu044.

Dong, Y. and Shen, S. (2018). Testing for rank invariance or similarity in program evaluation. *The Review of Economics and Statistics*, **100**, 78–85. doi: 10.1162/REST_a_00686.

Elwert, F. and Winship, C. (2014). Endogenous selection bias: the problem of conditioning on a collider variable. *Annual Review of Sociology*, **40**, 31–53.

England, P. *et al.* (2016). Do highly paid, highly skilled women experience the largest motherhood penalty? *American Sociological Review*, **81**, 1161–1189.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, **75**, 259–276.

Firpo, S., Fortin, N., and Lemieux, T. (2007). *Unconditional quantile regressions*. National Bureau of Economic Research #339. Cambridge.

Firpo, S., Fortin, N. M. and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, **77**, 953–973.

Firpo, S. and Pinto, C. (2016). Identification and estimation of distributional impacts of interventions using changes in inequality measures. *Journal of Applied Econometrics*, **31**, 457–486. doi: 10.1002/jae.2448.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Chapter 1 - decomposition methods in economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*. Vol. **4**. North Holland: Elsevier. pp. 1–102.

Frandsen, B. R. and Lefgren, L. J. (2018). Testing rank similarity. *The Review of Economics and Statistics*, **100**, 86–91. doi: 10.1162/REST_a_00675.

Frölich, M. and Melly, B. (2008). Estimation of quantile treatment effects with Stata. *Stata Journal*, **10**, 423–457.

Frölich, M. and Melly, B. (2010). Estimation of quantile treatment effects with Stata. *The Stata Journal*, **10**, 423–457.

Gangl, M. (2006). Scar effects of unemployment: an assessment of institutional complementarities. *American Sociological Review*, **71**, 986–1013.

Glauber, R. (2018). Trends in the motherhood wage penalty and fatherhood wage premium for low, middle, and high earners. *Demography*, **55**, 1663–1680.

Grätz, M. and Wiborg, N. (2020). Reinforcing at the top or compensating at the bottom? family background and academic performance in Germany, Norway, and the United States. *European Sociological Review*, **36**, 381–394.

Haupt, A. and Nollmann, G. (2021). Payroll-exempt labor incomes increase inequality at the top. *Social Forces*, **101**, 694–719.

Havnes, T. and Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of Public Economics*, **127**, 100–114. doi: 10.1016/j.jpubeco.2014.04.007.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, **81**, 945–960.

Killewald, A. and Bearak, J. (2014). Is the motherhood penalty larger for low-wage women? A comment on quantile regression. *American Sociological Review*, **79**, 350–357. doi: 10.1177/0003122414524574.

Koenker, R. (2005). *Quantile Regression*. Vol. **38**. Cambridge: Cambridge University Press.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, **46**, 33–50.

Koenker, R. and Hallock, K. (2001). Quantile regression: an introduction. *Journal of Economic Perspectives*, **15**, 43–56.

Lin, K. -H. and Weiss, I. (2019). Immigration and the wage distribution in the United States. *Demography*, **56**, 2229–2252. doi: 10.1007/s13524-019-00828-9.

Lindqvist, E. and Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: evidence from the swedish enlistment. *American Economic Journal: Applied Economics*, **3**, 101–128. doi: 10.1257/app.3.1.101.

Loyalka, P. *et al.* (2019). Pay by design: Teacher performance pay design and the distribution of student achievement. *Journal of Labor Economics*, **37**, 621–662.

Ludwig, V. and Brüderl, J. (2018). Is there a male marital wage premium? New evidence from the United States. *American Sociological Review*, **83**, 744–770. doi: 10.1177/0003122418784909.

Lundberg, I., Johnson, R., and Stewart, B. M. (2020). Setting the target: precise estimands and the gap between theory and empirics.

Lundberg, I., Johnson, R. and Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects statis-tical evidence to theory. *American Sociological Review*, **86**, 532–565. doi: 10.1177/00031224211004187.

Machado, J. A. F. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, **20**, 445–465.

Melly, B. (2005a). Decomposition of differences in distribution using quantile regression. *Labour Economics*, **12**, 577–590.

Melly, B. (2005b). Public-private sector wage differentials in Germany: evidence from quantile regression. *Empirical Economics*, **30**, 505–520.

Melly, B., and Wüthrich, K. (2017). Local quantile treatment effects. In *Handbook of Quantile Regression*. Chapman and Hall/CRC, pp. 145–164.

Moreno-Colom, S. (2017). The gendered division of housework time: Analysis of time use by type and daily frequency of household tasks. *Time & Society*, **26**, 3–27.

Morgan, S. L., and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Porter, S. R. (2015). Quantile regression: analyzing changes in distributions instead of means. In *Higher Education: Handbook of Theory and Research*. Springer, pp. 335–381.

Powell, D. (2020). Quantile treatment effects in the presence of covariates. *Review of Economics and Statistics*, **102**, 994–1005.

Powell, D. (2022). Quantile regression with nonadditive fixed effects. *Empirical Economics*.

Rios-Avila, F. (2020). Recentered influence functions (RIFs) in Stata: RIF regression and RIF decomposition. *The Stata Journal*, **20**, 51–94. doi: 10.1177/1536867X20909690.

Rios-Avila, F. and Maroto, M. L. (2020). Moving beyond linear regression: implementing and interpreting quantile regression models with fixed effects. *SocArXiv*, doi: 10.31235/osf.io/znj38. preprint: not peer reviewed.

Rothe, C. (2009). *Unconditional Partial Effects of Binary Covariates*. Retrieved from https://EconPapers.repec.org/RePEc:tse:wpaper:22190

Rothe, C. (2012). Partial distributional policy effects. *Econometrica*, **80**, 2269–2301.

Słoczyński, T. (2022). Interpreting OLS estimands when treatment effects are heterogeneous: smaller groups get larger weights. *The Review of Economics and Statistics*, **104**, 501–509. doi: 10.1162/rest_a_00953.

Wenz, S. E. (2019). What quantile regression does and doesn't do: a commentary on petscher and logan (2014). *Child Development*, **90**, 1442–1452.

Ying, Y. -H., Wu, C. -C. and Chang, K. (2013). The effectiveness of drinking and driving policies for different alcohol-related fatalities: a quantile regression analysis. *International Journal of Environmental Research and Public Health*, **10**, 4628–4644.